

UNDERSTANDING HUMAN-AI TRUST IN MILITARY AI APPLICATIONS: A MULTISTAGE HUMAN FACTORS FRAMEWORK

Kangwon Kim*

Department of Artificial Intelligence and Systems Science
Korea Army Academy at Yeongcheon
Yeongcheon, Republic of Korea

*Corresponding author's e-mail: alphatk21@kaay.ac.kr

Artificial intelligence (AI) is increasingly used in military systems to support intelligence, surveillance, reconnaissance, autonomous operations, logistics, command and control, and decision support. Effective use, however, depends not only on technical performance but also on how operators perceive, interpret, evaluate, and rely on AI outputs under uncertainty, time pressure, and adversarial interference. This study presents a conceptual integrative review informed by a structured supplementary literature search and synthesis. It synthesizes human factors, human–automation trust, human–AI teaming, military AI, and industrial engineering literature to develop an iterative multistage framework of human–AI trust in military applications. Trust is conceptualized as dynamic situational trust in a specific military AI system during task interaction, while dispositional, initial, and learned trust are treated as antecedent conditions. The framework distinguishes four analytically separable but functionally interdependent stages—perception, interpretation, evaluation, and trust judgment and reliance decision—and identifies four military-relevant moderators: system transparency, explainability, and behavioral predictability; data quality and information uncertainty; adversarial vulnerability; and operational and organizational context. The framework clarifies conceptual pathways to calibrated reliance, overtrust, and undertrust, and derives implications for task analysis, interface evaluation, human-centered design, operator training, and empirical validation.

Keywords: Human–AI Trust; Military AI Systems; Trust Calibration; Human–AI Teaming; Human Factors; Safety-Critical Systems

(Received on January 6, 2026; Accepted on August 2, 2026)

1. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) technologies are increasingly applied to military functions such as intelligence, surveillance, and reconnaissance (ISR), autonomous systems, cyber defense, logistics, command and control (C2), and decision support (Rashid *et al.*, 2023; Svenmarck *et al.*, 2018). In multi-domain operations (MDO), AI is also being considered to support rapid data analysis, mission planning, predictive targeting, and systems maintenance; however, these applications are expected to augment rather than fully replace human decision-makers, making effective human–AI teaming a central design concern (National Academies of Sciences, Engineering, and Medicine [NASEM], 2021). Related work in industrial engineering has also examined military manned–unmanned teaming from an optimization perspective, including location-routing and reconnaissance planning for combined manned and unmanned systems (Lee, 2025). The present review complements this application- and operations-oriented literature by focusing on the human factors dimension of military AI use: how operators perceive, interpret, evaluate, and rely on AI-enabled systems under uncertainty.

Trust is central to this human factors problem because reliance decisions are rarely determined by objective system performance alone. Prior human–automation studies indicate that trust, reliability disclosure, performance feedback, and explanations of aid errors can shape reliance on automated aids in military-relevant detection and combat identification tasks (Dzindolet *et al.*, 2003; Wang *et al.*, 2009). In a simulated combat identification task, for example, disclosing aid reliability helped response bias vary more appropriately with aid reliability, and trust in aid feedback was correlated with reliance on that feedback (Wang *et al.*, 2009). In a laboratory soldier-detection task, participants initially viewed an automated decision aid as trustworthy, but observing aid errors reduced trust and promoted disuse; providing an explanation of why the aid might err increased trust and reliance, even when that trust was not warranted (Dzindolet *et al.*, 2003). More broadly, trust in automation has been linked to appropriate reliance, misuse, and disuse in safety-critical systems (Lee & See, 2004; Parasuraman & Riley, 1997). The issue, therefore, is not whether operators should trust AI more, but whether their reliance is calibrated to available evidence about system capability and reliability. Accordingly, this review does not assume that mission effectiveness is determined by trust alone. Rather, it treats calibrated reliance as a human factors condition that can

influence safety-critical and mission-relevant performance when AI recommendations are embedded in operational decision-making.

Adjacent evidence from team and human–robot trust research broadens this reliance-focused argument while also clarifying its limits. A meta-analysis of intrateam trust found that trust is positively associated with team performance, with the relationship varying by task interdependence, authority differentiation, and skill differentiation (De Jong *et al.*, 2016). Human–robot trust research similarly indicates that perceived trust is shaped by human-, robot/system-, and environment-related factors rather than by technical performance alone (Hancock *et al.*, 2011). These findings do not establish a direct causal link between military AI trust and mission effectiveness; rather, they provide grounds for treating calibrated reliance as a multilevel human factors issue involving individual judgment, team coordination, system properties, and operational context.

Despite a substantial literature on human–automation, human–autonomy, and AI systems, trust research relevant to military AI remains fragmented in at least four ways. First, definitions and closely related constructs vary across interpersonal and organizational trust, automation trust, AI trust, trustworthiness, reliance, compliance, and warranted or unwarranted trust (Afroogh *et al.*, 2024; Jacovi *et al.*, 2021; Lee & See, 2004; Mayer *et al.*, 1995). Second, measurement approaches range from self-report trust scales and behavioral reliance indices to task performance, monitoring behavior, neurophysiological indicators, and team-process measures (Oh *et al.*, 2025; Schaefer *et al.*, 2016; Ulfert *et al.*, 2024). Third, levels of analysis differ: some studies examine individual cognition, others focus on human–AI or human–autonomy teams, and still others emphasize organizational, governance, or command-context conditions surrounding AI use (Fulmer & Gelfand, 2012; McFarland, 2022; Ulfert *et al.*, 2024). Fourth, application domains differ substantially: much empirical AI trust research has been conducted in civilian, commercial, laboratory, or otherwise non-adversarial settings, whereas military AI use involves contested information, adversarial manipulation, compressed decision time, and accountability constraints that are not fully captured in those settings (Glikson & Woolley, 2020; Henrique & Santos Jr., 2024; Mayer, 2023; Svenmarck *et al.*, 2018). These forms of fragmentation motivate an integrative framework that clarifies the focal trust construct, maps evidence across analytical levels, and specifies how military-relevant conditions enter trust-processing mechanisms.

Trust in military AI differs from trust in many civilian AI contexts not merely because the potential consequences are more severe, but because the mechanisms that shape trust operate under distinct technical and organizational constraints. Military AI systems can be deployed in contested information environments where data are sparse, delayed, degraded, compartmentalized, or intentionally manipulated. They may also be exposed to deliberate adversaries who exploit sensor, cyber, data, or model vulnerabilities (Svenmarck *et al.*, 2018). Decision cycles can be compressed, limiting the time available for independent verification and increasing the importance of timely but calibrated reliance (Montgomery, 2019; NASEM, 2021). The consequences of miscalibration are asymmetric: overtrust can contribute to automation bias, deceptive cue acceptance, misidentification, friendly-fire or targeting risk, collateral damage, or escalation, whereas undertrust can slow decision tempo, increase workload, or prevent use of otherwise useful AI capabilities (Lee & See, 2004; Mayer, 2023; Parasuraman & Riley, 1997; Wang *et al.*, 2009). These reliance decisions are further embedded in doctrine, rules of engagement, command authority, human-control requirements, governance arrangements, joint or coalition coordination requirements, and accountability structures (Marijan, 2025; McFarland, 2022). Thus, the relevant question is not simply whether military operators trust AI, but how technical uncertainty, adversarial pressure, and organizational constraints shape calibrated reliance during operational decision-making.

This concern is echoed in defense-oriented analyses, which treat trust as a central but often under-specified and under-operationalized issue in military human–machine teaming and AI adoption (Konaev & Chahal, 2021; Marijan, 2025). These analyses emphasize that technical trustworthiness alone is insufficient; calibrated reliance also depends on human factors, organizational practices, governance arrangements, and accountability structures that shape how AI systems are interpreted and used in operational settings.

The objective of this paper is to synthesize these fragmented perspectives into a multistage human factors framework for military AI trust. The framework focuses on dynamic situational trust in a specific military AI system during task interaction. It conceptualizes trust processing in terms of perception, interpretation, evaluation, and trust judgment and reliance decision, while treating these stages as analytically distinguishable rather than strictly linear. The framework further identifies four military-relevant moderators—(1) system transparency, explainability, and behavioral predictability; (2) data quality and information uncertainty; (3) adversarial vulnerability; and (4) operational and organizational context—that shape trust processing across stages. The theoretical contribution lies in process–condition integration: the framework links cognitive trust processing with technical uncertainty, adversarial interference, and military organizational conditions to clarify how calibrated reliance, overtrust, and undertrust can emerge in military AI applications. The practical contribution is to derive design and evaluation considerations for task analysis, interface evaluation, operator training, measurement, and human-centered design in industrial engineering and human factors practice.

2. REVIEW METHODOLOGY

2.1. Review design and supplementary literature review

This study uses a conceptual integrative review design informed by a structured supplementary literature search and synthesis. It was not designed as a systematic review or meta-analysis. Accordingly, the purpose is not to estimate aggregate effect sizes or exhaustively catalog all published studies, but to integrate theoretical, empirical, technical, military, and operational perspectives into a coherent framework for understanding trust calibration in military AI. This methodological positioning is consistent with guidance that literature reviews can serve as a research methodology when they synthesize fragmented interdisciplinary knowledge and support theory development (Snyder, 2019). It also follows the principles of integrative reviews, which allow empirical and theoretical sources using diverse methods to be combined to provide a more comprehensive understanding of a phenomenon (Whittemore & Knafl, 2005).

The structured supplementary search and synthesis were used to clarify the evidentiary basis of the conceptual synthesis, broaden the theoretical grounding of the framework, and make the literature selection and synthesis procedures explicit. The supplementary search focused on five bodies of literature: (1) foundational human–automation trust and reliance; (2) AI trust, explainability, and trustworthiness; (3) human–autonomy and human–AI teaming; (4) military AI applications, constraints, and governance; and (5) industrial engineering and human factors methods for system design, workload assessment, interface evaluation, and trust measurement.

2.2. Search strategy and inclusion criteria

Targeted searches were conducted using multidisciplinary indexing tools, publisher databases, and official repositories, including Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect, Taylor & Francis Online, SAGE Journals, Springer and Frontiers platforms, and official defense and National Academies sources. Because the purpose was conceptual synthesis rather than systematic enumeration, search strings were adapted to the syntax of each source while using consistent term blocks. The search strategy was designed to capture seminal human–automation literature from the 1990s onward and recent AI, human–AI teaming, and military AI literature published or available online at the time the supplementary search was completed.

Search terms were organized into four term blocks and combined across blocks: (1) trust and reliance terms, such as “trust in automation,” “human–AI trust,” “trust calibration,” “appropriate reliance,” “reliance,” “disuse,” and “misuse”; (2) AI and autonomy terms, such as “artificial intelligence,” “machine learning,” “autonomous systems,” “human–autonomy teaming,” “human–AI teaming,” “explainable AI,” “transparency,” and “predictability”; (3) military-context and technical-risk terms, such as “military AI,” “combat identification,” “manned–unmanned teaming,” “data quality,” and “adversarial machine learning”; and (4) human factors terms, such as “operator workload,” “situation awareness,” “team situation awareness,” “goal-directed task analysis,” “interface evaluation,” “training,” and “trust measurement.” Keyword searches were supplemented by backward and forward citation tracing from foundational trust, automation, situation awareness, human–AI teaming, and military AI sources. Seminal works and adjacent literature were retained when they clarified core constructs, levels of analysis, trust-related mechanisms, interface or display mechanisms, or industrial engineering implications directly relevant to the proposed framework.

Sources were included in the synthesis when they were conceptually relevant to the framework objectives and met at least one of the following criteria: they defined or modeled trust in automation, AI, autonomy, organizations, or teams; provided empirical evidence on trust, reliance, or performance in human–automation interaction; clarified transparency, explainability, data quality, adversarial vulnerability, workload, situation awareness, shared situation awareness, or mental models as mechanisms relevant to trust; examined military AI, human–autonomy teaming, human–AI teaming, or manned–unmanned teaming; or provided industrial engineering and human factors implications for system design, interface evaluation, training, workload assessment, or trust measurement. Sources were excluded when they focused only on technical AI performance without implications for human interaction or trust, addressed general civilian technology acceptance without transferable relevance to safety-critical, human–AI, human–automation, or military systems, or duplicated points already covered by more directly relevant studies. Peer-reviewed empirical, review, theoretical, and design-oriented studies were prioritized for claims about trust, reliance, situation awareness, interface design, and performance, whereas defense reports, policy analyses, theses, and application-oriented studies were used primarily to characterize military application contexts, governance issues, operational constraints, and industrial engineering relevance.

Titles, abstracts, and available full texts were screened for conceptual relevance to the framework objectives. The final synthesis was organized around sources that directly informed at least one component of the framework: the focal trust construct, the multistage trust-processing model, the military-relevant moderators, the moderator–stage mapping, or the industrial engineering and human factors implications for interface design, workload assessment, training, and measurement.

2.3. Conceptual synthesis and moderator selection

The synthesis proceeded in three steps. First, included sources were grouped by analytical level: operator cognition, system/interface properties, information environment, adversarial threat conditions, team processes, and operational and organizational context. Second, sources were mapped to trust-processing functions: perception of AI cues, interpretation through mental models, evaluation of capability, risk, and uncertainty, and trust judgment and reliance decision. Third, recurring cross-stage conditions were retained as framework moderators when they met three criteria: they appeared across multiple literature streams, shaped more than one trust-processing function, and had direct relevance to military AI applications or constraints. This approach distinguished moderator selection from simple topic frequency: a condition was retained only when it helped clarify how trust calibration or miscalibration could emerge across multiple stages of operator interaction with military AI. The four moderators are therefore not presented as an exhaustive taxonomy. They are conceptual categories identified through the supplementary literature search and synthesis because they appear across human factors, AI trust, and military AI sources and operate at complementary analytical levels. Together, they represent system/interface, information, threat, and operational and organizational levels of influence on trust processing.

3. CONCEPTUAL FOUNDATIONS OF HUMAN-AI TRUST

3.1. Trust definition, scope, and calibration

Trust is used in multiple ways across automation, AI, and team literatures. In this paper, trust is defined as an operator’s attitude or expectation that a specific AI system will help achieve task or mission goals under uncertainty and vulnerability, thereby shaping the willingness to rely on the system’s outputs or actions (Jacovi *et al.*, 2021; Lee & See, 2004; Mayer *et al.*, 1995). This definition distinguishes trust from trustworthiness, which refers to properties of the system or institution that make trust warranted, and from reliance, which refers to behavioral use or acceptance of system outputs. This distinction is important because trust can guide reliance without fully determining it; actual reliance can also be shaped by workload, time pressure, task demands, self-confidence, organizational rules, and other contextual constraints (Lee & See, 2004).

The focal construct of the framework is dynamic situational trust in a specific military AI system during task interaction. Dispositional trust, initial trust, learned trust, prior experience, and organizational culture are treated as antecedent or background conditions that shape the focal process rather than as the direct object of analysis. This distinction follows the three-layered model of trust in automation, which separates dispositional, situational, and learned factors influencing trust (Hoff & Bashir, 2015). It also helps distinguish the framework from broader organizational trust models and from team trust models that examine trust among multiple human and AI agents (Fulmer & Gelfand, 2012; Ulfert *et al.*, 2024).

Human–autonomy and human–AI teaming research further motivates this scope. Recent reviews and empirical work distinguish automation from autonomy and show that trust in AI-enabled teams is shaped by autonomy-as-teammate perceptions, shared mental models, team situation awareness, team cognition, and multilevel trust relationships (Andrews *et al.*, 2023; Endsley, 2023; Georganta & Ulfert, 2024; Lyons *et al.*, 2021; Schmutz *et al.*, 2024; Ulfert *et al.*, 2024).

Trust calibration is central to the framework. Appropriate trust occurs when reliance is aligned with system capability, task demands, uncertainty, and operational constraints. Miscalibration can take the form of overtrust, which contributes to misuse or automation bias, or undertrust, which contributes to disuse and unnecessary workload (Lee & See, 2004; Parasuraman & Riley, 1997). A recent neuroergonomic study of automation-trust measurement also highlights the relevance of examining whether human trust is aligned with automation capabilities, particularly in uncertain or urgent situations (Oh *et al.*, 2025).

Table 1 positions the proposed framework relative to representative prior trust frameworks and literature streams. The table is not intended to provide an exhaustive review of all trust models. Rather, it summarizes how selected foundational and adjacent literatures inform the present synthesis and clarifies what remains insufficiently specified for military AI trust. In particular, prior frameworks provide important foundations for trust definition, automation reliance, trust calibration, human–AI teaming, and military AI governance, whereas the present framework integrates these insights into a stage-sensitive and cross-level account of trust calibration and miscalibration in military AI applications.

Table 1. Representative prior trust frameworks and literature streams informing the proposed military AI trust framework

Framework or literature stream	Primary focus	Key contribution to this review	Unresolved issue addressed by the proposed framework
Mayer <i>et al.</i> (1995) and organizational/multilevel trust	Trustor–trustee vulnerability, ability,	Clarifies trust as vulnerability under	Not specific to AI, automation,

Framework or literature stream	Primary focus	Key contribution to this review	Unresolved issue addressed by the proposed framework
literature, including Fulmer and Gelfand (2012)	benevolence, integrity, risk, and levels of trust	uncertainty and distinguishes trust from antecedents, outcomes, and broader organizational conditions	adversarial environments, or military reliance decisions
Parasuraman and Riley (1997), Parasuraman <i>et al.</i> (2000), and Sheridan and Parasuraman (2005)	Automation use, misuse, disuse, abuse, and types/levels of automation	Provides functional decomposition of automation across information acquisition, analysis, decision/action selection, and action implementation	Does not specify military AI trust moderators or stage-sensitive trust calibration pathways
Lee and See (2004)	Trust in automation and appropriate reliance	Defines trust as an attitude that guides, but does not fully determine, reliance; emphasizes calibration, resolution, specificity, context, and display characteristics	Does not explicitly integrate adversarial manipulation, degraded data, or military organizational constraints
Hoff and Bashir (2015), Hancock <i>et al.</i> (2011), and Schaefer <i>et al.</i> (2016)	Empirical factors shaping trust in automation and human–robot/automation interaction	Clarifies dispositional, situational, and learned trust factors and identifies human-, system-, and environment-related antecedents of trust	Provides broad trust-factor evidence but not a military AI process–condition framework
Alarcon and Capiola (2025)	Information-processing model of human–AI/ML/XAI trust	Supports perception, interpretation, evaluation, and reliance-related processing in human–AI interaction	Not designed specifically around military adversarial conditions, degraded information environments, or multilevel operational constraints
Human–autonomy, human–AI teaming, and situation awareness literature, including Lyons <i>et al.</i> (2021), Andrews <i>et al.</i> (2023), Endsley (2023), Ulfert <i>et al.</i> (2024), Georganta and Ulfert (2024), and Schmutz <i>et al.</i> (2024)	Team cognition, shared mental models, situation awareness, AI as teammate, and multilevel team trust	Shows that trust is emergent, dynamic, embedded in team relationships, and linked to mental models and situation awareness	Often does not integrate technical and military moderators with specific trust-processing functions
Military AI and military HAT literature, including Svenmarck <i>et al.</i> (2018), NASEM (2021), Mayer (2023), McFarland (2022), and Marijan (2025)	Military AI applications, autonomy, trust–control tension, governance, data constraints, adversarial vulnerability, and operational constraints	Identifies military-relevant risks such as time pressure, accountability, degraded information, adversarial vulnerability, and human-control requirements	Often discusses conditions separately rather than integrating them into an operator-level trust process
Proposed framework	Iterative multistage trust processing in military AI	Integrates cognitive trust-processing stages with system/interface, information, threat, and operational and organizational moderators	Provides a stage-sensitive and cross-level basis for design, training, measurement, and empirical testing

3.2. Iterative multistage trust processing

The proposed process model uses four analytically distinguishable stages: perception, interpretation, evaluation, and trust judgment and reliance decision. Perception refers to the operator's intake of AI-generated cues, alerts, recommendations, confidence information, uncertainty displays, and interface signals. Interpretation refers to the assignment of meaning to those cues using mental models, task knowledge, prior experience, and assumptions about system behavior. Evaluation refers to the operator's assessment of AI output credibility, capability, risk, uncertainty, and consistency with other mission evidence. Trust judgment and reliance decision refer to the formation of a trust appraisal and the selection of a reliance response, such as acceptance, verification, conditional reliance, override, deferral, or disengagement.

These stages should not be interpreted as a rigidly sequential cognitive chain. Rather, they are functional distinctions that help analyze where trust-supporting or trust-disrupting mechanisms enter human-AI interaction. In real operations, the stages overlap, recursively update one another, and are shaped by feedback from system outcomes. This interpretation draws on information-processing accounts of human-AI/ML trust and aligns with human-automation models that distinguish information acquisition, information analysis, decision/action selection, and action implementation while recognizing that automation changes human activity and coordination demands (Alarcon & Capiola, 2025; Parasuraman *et al.*, 2000). It is also consistent with trust-in-automation research that treats trust and reliance as a dynamic interaction among the operator, context, automation, and interface, rather than as a static attitude-behavior link (Lee & See, 2004). In addition, situation-awareness research in human-AI teams supports the importance of perception, comprehension, mental models, projection, and shared displays for effective oversight and calibrated trust (Endsley, 2023). Human-automation interaction research similarly shows that inadequate feedback, misunderstanding of automation, and overreliance can interact across system use (Sheridan & Parasuraman, 2005).

The value of the staged framework is analytical rather than deterministic. It allows researchers and designers to ask whether trust miscalibration or reliance failure originated because relevant cues were not perceived, because available cues were misunderstood, because risk and uncertainty were evaluated incorrectly, or because a calibrated trust judgment did not translate into appropriate reliance behavior under operational constraints. It also allows multiple failure pathways to be represented. For example, a spoofed sensor input can affect perception, a misleading explanation can affect interpretation, hidden data bias can affect evaluation, and command pressure can affect the final reliance decision.

Trust trajectory refers to the pattern of change in an operator's trust judgments and reliance behavior across repeated interactions, observed system successes, failures, anomalies, explanations, and changing mission conditions. Human-automation research conceptualizes trust and reliance as a dynamic process in which system behavior, feedback, context, and prior experience influence subsequent reliance (Lee & See, 2004). Empirical work on automation also shows that operators' trust can be reduced by signs of automation incompetence and that trust can be related to use and monitoring behavior (Muir & Moray, 1996). Algorithm aversion research further shows that people can lose confidence in algorithms after observing errors even when algorithms outperform humans (Dietvorst *et al.*, 2015). In military AI, trust trajectories may therefore involve both justified recalibration and problematic disuse following salient failures.

Cognitive load, workload, and expertise are included as background conditions because military AI use often occurs under multitasking, time pressure, and competing attentional demands. Multiple resource theory explains why concurrent perceptual, cognitive, and response demands can interfere with operator performance when tasks compete for limited resources (Wickens, 2008). In AI-supported military tasks, workload can reduce the attention available for uncertainty cues, explanation processing, and independent verification. Expertise and prior experience can improve diagnostic assessment and mental models, but they can also introduce skepticism, automation disuse, or overconfidence in human judgment when operators discount useful AI support after salient errors or when prior doctrine conflicts with AI-generated recommendations (Dietvorst *et al.*, 2015; Kean, 2022; Lee & See, 2004). These factors are treated as background conditions rather than primary military-relevant moderators because they shape the operator's processing capacity and prior expectations rather than defining the system/interface, information, threat, or operational and organizational conditions represented in the framework.

4. MILITARY-RELEVANT MODERATORS OF TRUST

The supplementary literature search and synthesis identified four recurring moderator categories relevant to military AI trust. They were selected because they appeared across multiple literature streams and because each operates through a primary mechanism in the trust process. System transparency, explainability, and behavioral predictability concern what the AI system communicates and how consistently it behaves. Data quality and information uncertainty concern the input and evidence base from which AI outputs are produced. Adversarial vulnerability concerns deliberate manipulation of the AI system, data stream, model input, or operating environment. Operational and organizational context concerns the mission, doctrinal, command, joint or coalition coordination, and accountability structures that shape whether trust judgments can be translated into action.

These moderators are not independent causes of trust in a simple linear sense. They affect the availability, meaning, credibility, and actionability of cues used by operators. They can also interact. For example, poor transparency can prevent operators from noticing degraded data, adversarial manipulation can make otherwise familiar system behavior suspicious, and command doctrine can determine whether an operator is permitted to rely on an AI recommendation even when personal trust is high. Table 2 summarizes the moderator–stage mapping used in the proposed framework. The table is intended as a conceptual mapping rather than a causal ranking; it identifies the stages and feedback processes most directly affected by each moderator while recognizing that moderator effects can accumulate, interact, or offset one another across stages.

4.1. System transparency, explainability, and behavioral predictability

The system-level moderator combines three related but distinct constructs: transparency, explainability, and behavioral predictability. Transparency refers to the observability of system state, input sources, uncertainty, assumptions, limitations, and processing logic. Explainability refers to the availability of reasons or causal accounts that help operators understand why a specific output or recommendation was generated. Predictability refers to the operator's ability to anticipate how the AI will behave under similar conditions. These concepts overlap, but they influence trust through different mechanisms: transparency supports cue perception and situation awareness, explainability supports interpretation and mental-model formation, and predictability supports expectation formation and reliance planning (Barredo Arrieta *et al.*, 2020; Endsley, 2023; Miller, 2019).

More information is not always better. Excessive, poorly timed, or poorly integrated transparency can increase workload, distract attention, or reduce trust when the information does not support the operator's current goals and expectations (Endsley, 2023; Kizilcec, 2016). For military AI, the design problem is therefore not simply to maximize transparency, but to provide role-appropriate, mission-relevant, and time-sensitive information that supports calibrated assessment. For example, an ISR analyst may need access to data provenance and uncertainty estimates, whereas a time-constrained vehicle operator may need a concise confidence cue and an actionable explanation of why the AI changed its recommendation.

4.2. Data quality and information uncertainty

Data quality and information uncertainty affect trust indirectly by shaping the reliability, consistency, and diagnostic value of AI outputs. Military data are often sparse, delayed, degraded, compartmentalized, sensor-dependent, biased toward past operations, or collected under conditions that differ from the deployment environment. High-stakes AI literature shows that data quality problems can cascade downstream and create delayed or invisible effects on system outputs (Sambasivan *et al.*, 2021). AI safety literature also highlights distributional shift as a practical risk when deployment inputs differ from training conditions (Amodei *et al.*, 2016).

In military AI, data limitations can influence perception when uncertainty cues are unavailable or poorly displayed, interpretation when operators do not understand data provenance, evaluation when output reliability cannot be assessed, and reliance decisions when time pressure discourages verification. These conditions do not uniformly lower trust. If limitations are visible, operators may appropriately reduce reliance. If limitations are hidden, operators may maintain unwarranted trust. Svenmarck *et al.* (2018) identify lack of sufficient data as a major challenge for military AI, and NASEM (2021) emphasizes that AI systems for multi-domain operations must be reliable and robust across varied future missions while accounting for perceptual limitations and hidden biases.

4.3. Adversarial vulnerability

Adversarial vulnerability is a defining feature of military AI trust. Unlike many non-adversarial civilian AI settings, military environments involve opponents who may intentionally manipulate sensors, communication channels, training data, model inputs, or contextual cues. Technical work on adversarial examples shows that small, intentionally crafted perturbations can induce high-confidence misclassification in machine learning models (Goodfellow *et al.*, 2014). Transferability research further shows that black-box attacks can be mounted even when the attacker has limited information about the target model (Papernot *et al.*, 2016).

The trust mechanism is not that adversarial threats always reduce trust. Rather, adversarial vulnerability disrupts calibration by making it difficult for operators to determine whether an AI output reflects system competence, data degradation, environmental ambiguity, or deliberate manipulation. This uncertainty can produce both overtrust and undertrust. Overtrust occurs when manipulated outputs remain plausible and are accepted without verification. Undertrust occurs when operators become broadly skeptical of AI outputs after detecting or suspecting manipulation. In both cases, the central issue is not the direction of trust change but the loss of diagnostic clarity in trust assessment.

4.4. Operational and organizational context

Operational and organizational context moderates how trust judgments become behavior. Military operators do not rely on AI in isolation; they act within doctrine, rules of engagement, command authority, mission tempo, joint or coalition coordination, classification restrictions, and accountability regimes. Mayer (2023) shows that trust calibration issues differ across military AI applications such as data integration and analysis, autonomous systems, and decision support. McFarland (2022) further emphasizes that military AI development is often framed in terms of trust, while legal and ethical discussions emphasize human control. These paradigms can pull design and governance in different directions.

Distinct contextual mechanisms should be differentiated. Time pressure reduces opportunities for verification and pushes operators toward rapid reliance or rejection. Doctrine and rules of engagement shape whether a trust judgment can be translated into authorized action. Command authority may create a gap between individual trust and organizationally permitted use. Joint or coalition coordination can affect source credibility, data sharing, interoperability, and classification. Organizational culture can normalize either excessive skepticism or uncritical automation use. Montgomery (2019) highlights that compressed reaction times and gray-zone ambiguity make predictability, transparency, traceability, and accountability central to trust in military AI systems.

Table 2. Moderator–stage mapping in the proposed military AI trust framework

Moderator	Analytical level	Primary trust mechanism	Stages/processes most directly affected
System transparency, explainability, and behavioral predictability	System/interface behavior	Makes system state, rationale, uncertainty, limitations, and expected behavior available to operators	Perception, interpretation, evaluation, trust judgment, and reliance decision
Data quality and information uncertainty	Input and evidence environment	Shapes output reliability, cue diagnosticity, confidence, and uncertainty assessment	Perception, interpretation, evaluation
Adversarial vulnerability	Threat–system interaction	Creates uncertainty about whether outputs reflect competence, data degradation, environmental ambiguity, or deliberate manipulation	Perception, interpretation, evaluation, trust judgment, and reliance decision
Operational and organizational context	Mission, command, doctrine, joint or coalition coordination, and accountability level	Determines how trust judgments are weighted and translated into authorized reliance behavior	Evaluation, trust judgment, and reliance decision, feedback, and calibration

Note. The mapping is conceptual rather than exclusive. Representative supporting literature for each moderator is discussed in Sections 4.1–4.4. Moderator effects can accumulate, interact, or offset one another across stages.

5. INTEGRATED MULTISTAGE FRAMEWORK

Figure 1 presents the integrated framework. The center of the figure represents operator-level trust processing: perception of AI cues, interpretation using mental models, evaluation of capability, risk, and uncertainty, and trust judgment and reliance decision. The upper layer represents system/interface, information, threat, and operational/organizational moderators. The lower layer represents reliance behavior, mission outcome, and feedback. The feedback loop indicates that trust is updated as operators observe AI successes, failures, anomalies, explanations, or mission consequences.

Figure 1 is intended to function as a mechanism-oriented architecture rather than a simple input–process–output pathway. Military AI capabilities and mission outcomes remain important, but the figure centers on the intermediate trust-processing functions through which technical, informational, adversarial, operational, and organizational conditions shape operator reliance behavior. Accordingly, the figure should be read as an iterative multilevel model: moderators shape the cues available to operators, operators process those cues through overlapping cognitive functions, reliance behavior produces operational consequences, and feedback updates subsequent trust judgments.

Military AI application domains are not depicted as separate inputs in Figure 1 because the framework is intended to describe a cross-application trust mechanism. Domains such as ISR, autonomous platforms, cyber defense, C2, and targeting are treated as operational contexts in which the moderator layer operates. Similarly, adoption, rejection, override, and human–AI teaming are represented through trust judgment, reliance decision, and reliance behavior, whereas mission-relevant

performance indicators such as ISR quality, C2 tempo, decision accuracy, and targeting risk are treated as downstream mission outcomes and evaluation criteria developed in Section 6.

The framework combines multiple levels of analysis by specifying how they interact. System/interface-level conditions shape what cues are available to the operator and how those cues are displayed. Information-level conditions shape the reliability and diagnostic value of AI outputs. Threat-level conditions shape whether cues can be trusted as authentic or unmanipulated. Operational and organizational conditions shape how cues are interpreted, weighted, and translated into reliance behavior. The operator-level trust process is therefore embedded within broader sociotechnical conditions rather than isolated from them.

The contribution lies in integrating established insights about dynamic trust, transparency, and context into a stage-sensitive and cross-level account of military AI trust. Prior human–automation and human–AI research has already shown that trust is dynamic, that transparency and explainability shape system understanding, and that context affects reliance (Endsley, 2023; Lee & See, 2004). The proposed framework extends these insights by mapping military-relevant system/interface, information, threat, and operational/organizational conditions onto trust-processing stages. In doing so, the framework clarifies conceptual pathways through which calibrated reliance, overtrust, and undertrust can emerge in military AI applications. It also helps explain why the same AI performance profile can produce different reliance outcomes across mission types, interface designs, operator expertise levels, and command environments.

For example, an AI-supported targeting system may produce a high-confidence recommendation based on degraded sensor data. If data quality limitations are not visible, the operator may perceive the recommendation as reliable and overtrust the output. If uncertainty is visible but not interpretable under time pressure, the operator may reject the recommendation even when it remains useful. If adversarial manipulation is suspected, even valid outputs may be discounted. If command doctrine requires human verification, personal trust may not translate into immediate action. These pathways illustrate why military AI trust should be analyzed as a multistage and multilevel process rather than a single attitude or outcome.

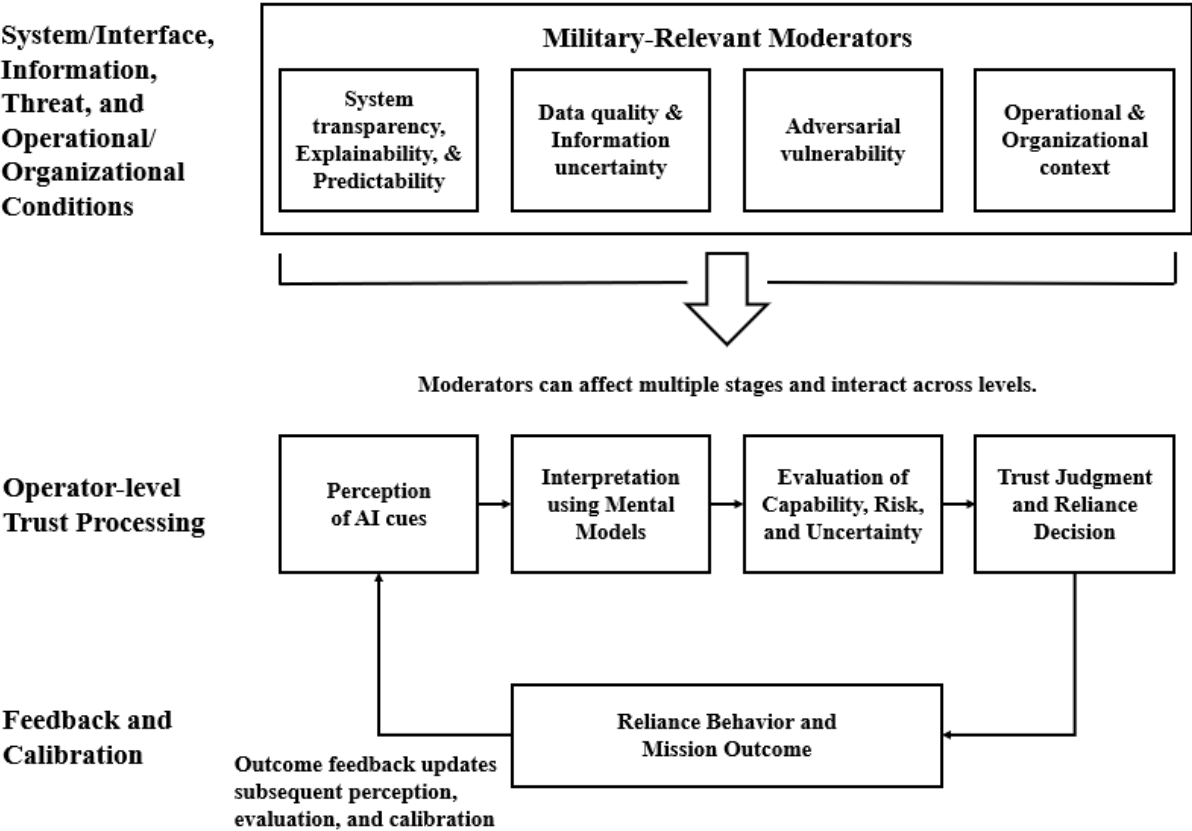


Figure 1. Integrated multistage framework of human–AI trust in military applications

6. IMPLICATIONS FOR INDUSTRIAL ENGINEERING AND HUMAN FACTORS

The framework is intended to inform industrial engineering, human factors, and ergonomics practice by translating trust-processing concepts into candidate design, training, and evaluation considerations. It can structure task analysis by identifying which task stages involve AI cue detection, explanation comprehension, uncertainty assessment, verification, and reliance decisions. It can also guide interface-evaluation questions about what information may need to be visible to operators, when it should be presented, and how it should be integrated with ongoing task demands. This orientation is consistent with situation-awareness-oriented design approaches that derive display requirements from goals, decisions, and situation-awareness needs and evaluate displays using situation awareness, workload, and performance measures (Endsley, 2023). For human-centered design, the framework identifies potential trust vulnerabilities that can be translated into candidate requirements for uncertainty visualization, explanation timing, provenance displays, workload management, and escalation procedures. For operator training, it points to cognitive skills that may be needed for calibrated reliance under degraded, time-pressured, or adversarial conditions.

This contribution complements optimization-oriented industrial engineering research on military manned–unmanned teaming. Location-routing and reconnaissance planning models provide tools for modeling allocation and coordination problems in combined manned and unmanned systems (Lee, 2025), but operators still need interfaces, procedures, and training that help them assess when reliance on AI-generated recommendations is appropriate. Similarly, a recent neuroergonomic study of automation-trust measurement illustrates how trust and mistrust states may be examined using cognitive indicators, providing a possible empirical direction for evaluating trust calibration in uncertain or urgent automation contexts (Oh *et al.*, 2025). Adjacent safety-critical human–AI teamwork literature in aviation also emphasizes role clarity, task allocation, situation awareness, training, and overreliance management, which are consistent with the industrial engineering-oriented considerations proposed here (Korentsides *et al.*, 2026).

Table 3 translates the framework into illustrative design, training, and measurement considerations. These examples are not intended as exhaustive or empirically validated prescriptions. Their purpose is to show how the conceptual framework can be used to formulate practitioner-facing questions in task analysis, interface evaluation, training design, and empirical validation.

Table 3. Illustrative design, training, and evaluation considerations derived from the framework

Trust-processing issue	Candidate design or training consideration	Example evaluation measure
Relevant uncertainty cue is not noticed	Use salient but workload-sensitive uncertainty visualization and alert prioritization	Cue detection rate; fixation or dwell time on uncertainty cues; missed uncertainty alerts
AI explanation is misunderstood	Provide role-specific explanations tied to mission questions rather than technical model details	Explanation comprehension; response accuracy; confidence calibration
Operator cannot compare AI confidence with mission evidence	Display confidence, data provenance, and evidence conflicts in an integrated view	Calibration error; confidence–evidence agreement; verification accuracy
Overreliance on high-confidence recommendation	Introduce risk-calibrated verification prompts or second-source checks for high-risk decisions	Verification frequency; automation-bias indicators; false acceptance rate
Underreliance after salient AI error	Train operators on failure modes, reliability ranges, and recovery after anomalies	Appropriate reliance rate; post-error reliance calibration; trust recovery curve; disuse frequency
Suspicion of adversarial manipulation	Provide provenance, integrity, anomaly, and model-input consistency indicators	Anomaly recognition; time to detect spoofing; response appropriateness
Time pressure limits evaluation	Use tiered displays that provide a concise first layer and drill-down explanation when time permits	Decision time; workload; accuracy under time pressure
Cognitive workload overwhelms monitoring	Apply workload-sensitive alerting and distribute information across	Subjective workload rating; dual-task performance; monitoring failures

Trust-processing issue	Candidate design or training consideration	Example evaluation measure
	perceptual modalities where appropriate	
Organizational policy constrains reliance	Align interface prompts and training scenarios with doctrine, rules of engagement, and command procedures	Procedure compliance; scenario-based reliance decisions; after-action review

Note. The considerations are illustrative rather than prescriptive. Specific design choices and evaluation measures should be adapted to the mission, AI function, operator role, and organizational constraints.

7. RESEARCH AGENDA AND LIMITATIONS

The proposed framework suggests several empirical priorities. First, future studies should examine whether the four trust-processing stages are empirically distinguishable in military AI tasks or whether some stages collapse under time pressure. Behavioral measures, process tracing, eye tracking, communication analysis, and physiological measures could be combined to examine how operators move from cue perception to trust judgment, reliance decision, and observable reliance behavior.

Second, future work should examine stage-specific and cross-stage effects of moderators. For example, experiments could manipulate transparency, explanation format, data provenance, or adversarial uncertainty to test whether effects occur primarily through perception, interpretation, evaluation, trust judgment and reliance decision, or downstream reliance behavior. Studies should also examine interactions among moderators, such as whether transparency mitigates the effect of poor data quality or whether adversarial suspicion reduces the value of explanation.

Third, trust trajectories should be studied longitudinally. Many trust studies involve short laboratory interactions, but military AI trust is likely shaped by repeated exposure, training, mission feedback, observed failures, and organizational norms. Longitudinal and simulation-based studies can examine trust decay, trust repair, algorithm aversion, and conditions under which operators regain calibrated reliance after AI errors (Dietvorst *et al.*, 2015).

Fourth, research should connect trust measurement with industrial engineering and human factors practice. Measures should include self-reported trust, reliance behavior, verification behavior, workload, situation awareness, task performance, and potentially neuroergonomic indicators (Endsley, 2023; Oh *et al.*, 2025). Evaluation protocols should be designed around mission-relevant scenarios rather than generic AI use cases, particularly for contested, degraded, or time-compressed conditions.

Several limitations remain. This paper provides a conceptual integrative framework rather than an empirical validation or meta-analysis. The structured supplementary literature search and synthesis was transparent but not exhaustive. Some military AI sources are technical reports, theses, or policy analyses rather than peer-reviewed empirical studies. The framework also emphasizes operator-level trust processing while only partially addressing broader institutional, coalition, and interstate trust issues. Accordingly, the framework should be treated as a testable conceptual model rather than as a validated predictive model. Future research should therefore validate the framework across military domains, compare it with civilian safety-critical AI systems, and examine how organizational and coalition-level trust conditions interact with individual operator cognition.

8. CONCLUSION

This review develops an integrated framework for understanding human–AI trust in military applications by clarifying the review methodology, defining the focal trust construct, comparing prior trust frameworks, and specifying how military-relevant moderators influence trust-processing stages. Human–AI trust in military applications is conceptualized as dynamic situational trust that develops through iterative perception, interpretation, evaluation, and trust judgment and reliance decision, rather than as a rigid sequence or static attitude.

The framework’s theoretical contribution lies in integrating cognitive trust processing with system transparency, explainability, and behavioral predictability; data quality and information uncertainty; adversarial vulnerability; and operational and organizational context. This integration clarifies why trust miscalibration in military AI cannot be understood from system accuracy, transparency, or operator disposition alone. Instead, calibrated reliance is shaped by interactions among operator cognition, AI system behavior, information conditions, threat environments, and military command structures.

For industrial engineering and human factors, the framework provides a conceptual basis for task analysis, interface evaluation, human-centered design, operator training, and empirical measurement. By identifying where trust may fail and which moderators may shape those failures, the framework can inform efforts to design and evaluate military AI systems not

only for technical capability, but also for understandability, evaluability, governability, and conditions for appropriate reliance in safety-critical operational contexts.

REFERENCES

- Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 11(1), 1568. DOI: <https://doi.org/10.1057/s41599-024-04044-8>
- Alarcon, G. M., & Capiola, A. (2025). Explicating the trust process for effective human interaction with artificial intelligence and machine learning systems. *Frontiers in Computer Science*, 7, 1662185. DOI: <https://doi.org/10.3389/fcomp.2025.1662185>
- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.
- Andrews, R. W., Lilly, J. M., Srivastava, D., & Feigh, K. M. (2023). The role of shared mental models in human–AI teams: A theoretical review. *Theoretical Issues in Ergonomics Science*, 24(2), 129-175. DOI: <https://doi.org/10.1080/1463922X.2022.2061080>
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. DOI: <https://doi.org/10.1016/j.inffus.2019.12.012>
- De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, 101(8), 1134-1150. DOI: <https://doi.org/10.1037/apl0000110>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114-126. DOI: <https://doi.org/10.1037/xge0000033>
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human–Computer Studies*, 58(6), 697-718. DOI: [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)
- Endsley, M. R. (2023). Supporting human–AI teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, 140, 107574. DOI: <https://doi.org/10.1016/j.chb.2022.107574>
- Fulmer, C. A., & Gelfand, M. J. (2012). At what level and in whom we trust: Trust across multiple organizational levels. *Journal of Management*, 38(4), 1167-1230. DOI: <https://doi.org/10.1177/0149206312439327>
- Georganta, E., & Ulfert, A.-S. (2024). Would you trust an AI team member? Team trust in human–AI teams. *Journal of Occupational and Organizational Psychology*, 97(4), 1212-1241. DOI: <https://doi.org/10.1111/joop.12504>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627-660. DOI: <https://doi.org/10.5465/annals.2018.0057>
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human–robot interaction. *Human Factors*, 53(5), 517-527. DOI: <https://doi.org/10.1177/0018720811417254>
- Henrique, B. M., & Santos, E., Jr. (2024). Trust in artificial intelligence: Literature review and main path analysis. *Computers in Human Behavior: Artificial Humans*, 2(1), 100043. DOI: <https://doi.org/10.1016/j.chbah.2024.100043>

- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407-434. DOI: <https://doi.org/10.1177/0018720814547570>
- Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 624-635). DOI: <https://doi.org/10.1145/3442188.3445923>
- Kean, C. R. (2022). *Trust and human–autonomy teaming* [Master's thesis, U.S. Army Command and General Staff College].
- Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390-2395). DOI: <https://doi.org/10.1145/2858036.2858402>
- Konaev, M., & Chahal, H. (2021, February 18). *Building trust in human–machine teams*. Brookings Institution.
- Korentsides, J., Merwin, E. R., Berger, L., Laskey, L., Winter, S. R., Sobel, B., & Keebler, J. R. (2026). The use of artificial intelligence (AI) in the flight deck: Enhancing human–AI teamwork in aviation. *Journal of the Air Transport Research Society*, 6, 100099. DOI: <https://doi.org/10.1016/j.jatrs.2025.100099>
- Lee, J. (2025). Capacitated location-routing problem for a combined manned-unmanned teaming system using Lagrangian relaxation and location-based heuristic. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 32(5). DOI: <https://doi.org/10.23055/ijietap.2025.32.5.10347>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80. DOI: https://doi.org/10.1518/hfes.46.1.50_30392
- Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology*, 12, 589585. DOI: <https://doi.org/10.3389/fpsyg.2021.589585>
- Marijan, B. (2025). *Trust by design? AI in military applications* (Policy Brief No. 218). Center for International Governance Innovation.
- Mayer, M. (2023). Trusting machine intelligence: Artificial intelligence and human–autonomy teaming in military operations. *Defense & Security Analysis*, 39(4), 521-538. DOI: <https://doi.org/10.1080/14751798.2023.2264070>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734. DOI: <https://doi.org/10.5465/amr.1995.9508080335>
- McFarland, T. (2022). Reconciling trust and control in the military use of artificial intelligence. *International Journal of Law and Information Technology*, 30(4), 472-483. DOI: <https://doi.org/10.1093/ijlit/caad008>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38. DOI: <https://doi.org/10.1016/j.artint.2018.07.007>
- Montgomery, C. S. (2019). *Trust in the machine: AI, autonomy, and military decision making with lethal consequences*. U.S. Naval War College.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II: Experimental studies of trust and intervention in a process control simulation. *Ergonomics*, 39(3), 429-460. DOI: <https://doi.org/10.1080/00140139608964474>
- National Academies of Sciences, Engineering, and Medicine. (2021). *Human–AI teaming: State-of-the-art and research needs*. The National Academies Press. DOI: <https://doi.org/10.17226/26355>
- Oh, S., Seong, Y., Yi, S., & Park, S. (2025). Topographic maps of brain regions to measure human trust in automation. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 32(5), 1381-1395. DOI: <https://doi.org/10.23055/ijietap.2025.32.5.10401>

- Papernot, N., McDaniel, P., & Goodfellow, I. (2016). Transferability in machine learning: From phenomena to black-box attacks using adversarial samples. *arXiv preprint arXiv:1605.07277*.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253. DOI: <https://doi.org/10.1518/001872097778543886>
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(3), 286-297. DOI: <https://doi.org/10.1109/3468.844354>
- Rashid, A. B., Kausik, A. K., Sunny, A. A. H., & Bappy, M. H. (2023). Artificial intelligence in the military: An overview of the capabilities, applications, and challenges. *International Journal of Intelligent Systems*, 2023(1), Article 8676366. DOI: <https://doi.org/10.1155/2023/8676366>
- Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. (2021). “Everyone wants to do the model work, not the data work”: Data cascades in high-stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). DOI: <https://doi.org/10.1145/3411764.3445518>
- Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, 58(3), 377-400. DOI: <https://doi.org/10.1177/0018720816634228>
- Schmutz, J. B., Outland, N., Kerstan, S., Georganta, E., & Ulfert, A.-S. (2024). AI-teaming: Redefining collaboration in the digital era. *Current Opinion in Psychology*, 58, 101837. DOI: <https://doi.org/10.1016/j.copsyc.2024.101837>
- Sheridan, T. B., & Parasuraman, R. (2005). Human–automation interaction. *Reviews of Human Factors and Ergonomics*, 1(1), 89-129. DOI: <https://doi.org/10.1518/155723405783703082>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. DOI: <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Svenmarck, P., Luotsinen, L., Nilsson, M., & Schubert, J. (2018, May). Possibilities and challenges for artificial intelligence in military applications. In *Proceedings of the NATO big data and artificial intelligence for military decision making specialists' meeting* (Vol. 1).
- Ulfert, A.-S., Georganta, E., Centeio Jorge, C., Mehrotra, S., & Tielman, M. (2024). Shaping a multidisciplinary understanding of team trust in human–AI teams: A theoretical framework. *European Journal of Work and Organizational Psychology*, 33(2), 158-171. DOI: <https://doi.org/10.1080/1359432X.2023.2200172>
- Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors*, 51(3), 281-291. DOI: <https://doi.org/10.1177/0018720809338842>
- Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52(5), 546-553. DOI: <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449-455. DOI: <https://doi.org/10.1518/001872008X288394>